

## ABSTRACT

A language input architecture converts input strings of phonetic text (e.g., Chinese Pinyin) to an output string of language text (e.g., Chinese Hanzi) in a manner that minimizes typographical errors and conversion errors that occur during conversion from the phonetic text to the language text. The language input architecture has a search engine, one or more typing models, a language model, and one or more lexicons for different languages. Each typing model is trained on real data, and learns probabilities of typing errors. The typing model is configured to generate a list of probable typing candidates that may be substituted for the input string based on probabilities of how likely each of the candidate strings was incorrectly entered as the input string. The probable typing candidates may be stored in a database. The language model provides probable conversion strings for each of the typing candidates based on probabilities of how likely a probable conversion output string represents the candidate string. The search engine combines the probabilities of the typing and language models to find the most probable conversion string that represents a converted form of the input string. By generating typing candidates and then using the associated conversion strings to replace the input string, the architecture eliminates many common typographical errors. When multiple typing models are employed, the architecture can automatically distinguish among multiple languages without requiring mode switching for entry of the different languages.